The Report Committee for Sibei Wen
Certifies that this is the approved version of the following report:

ESTIMATION OF MULTIPLE MEDIATOR MODEL

APPROVED BY
SUPERVISING COMMITTEE:

S.Natasha Beretvas, Supervisor

Daniel A. Powers

ESTIMATION OF MULTIPLE MEDIATOR MODEL

by

Sibei Wen, B.Eco.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2013

Abstract

ESTIMATION OF MULTIPLE MEDIATOR MODEL

Sibei Wen, M.S. Stat

The University of Texas at Austin, 2013

Supervisor: S.Natasha Beretvas

Models for mediation are widely used in psychology, behavior science and education because they help researchers understand how a causal effect happens through one or several mediating variables. And more complex mediation models that incorporate multiple mediators are increasingly being assessed. This report uses a generated dataset to provide an overview of the assessment of direct effects and indirect effects in multiple mediator models. Use of a multiple comparison-based procedure for testing a set of hypotheses simultaneously while controlling the experiment-wise type I error rate is used to calculate a confidence interval for each pairwise contrast of mediated effects. Three approaches will be used to test hypotheses concerning the contrast between pairs of mediator effects. These approaches include 1) an assumption of zero covariance between parameters from different models, 2) assumption of a non-zero covariance between parameters from different models and 3) use of bootstrapping. Results are provided and discussed.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Finding support for relationships between variables is a fundamental contribution of statistics. A researcher might find a correlation between two related variables. But the existence of a correlation is a necessary, not a sufficient condition, which cannot explain why or how this causal effect might happen between the two variables. Rarely are researchers satisfied with detection of only a correlation. Instead, researchers are typically more interested in understanding how such an effect comes to be. Thus, this invokes the idea of a mediation model. Mediation model is important in identifying and explicating the process that underlies an observed relationship between a response variable and a dependent variable through the consideration of one or more explanatory variables, which are also known as mediator variables. In other words, a mediator model is such a kind of model that involves one or multiple variables which are connected in some form of "process" relationship. The primary role of mediator model is that, once a relationship between an independent and dependent variable is found, mediation analysis can help investigate the processes underlying the observed relationship.

The mediator model is widely applied in social, educational and behavioral science fields, because researchers from these fields are more interested in the process through which some treatment might have an effect rather than just that the treatment has an effect. For example, an experiment might be intended to improve students' achievement by improving teaching methods. The results might indicate that there is a strong relationship between the teaching methods and students' achievement. But the researcher might be more interested in how the teaching methods influence students' achievement. For instance, the effect might have occurred by the teaching methods having changed students' attitudes or by changing their behaviors, or both. To solve this problem, a researcher would need to use a mediator model.

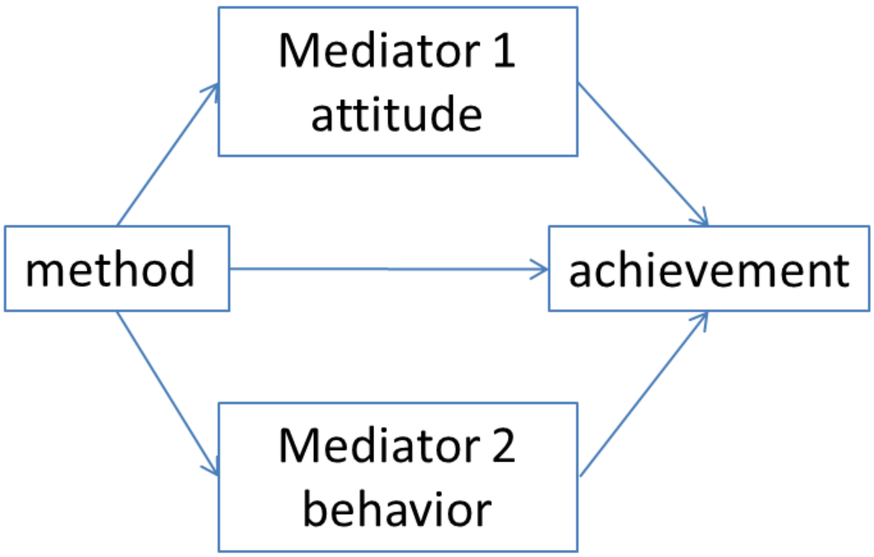

Figure 1: *Example of Teaching Method affects Students' Achievements through mediators like Students' Attitudes and Student's Behaviors.*

In the case mentioned above, teaching methods would constitute the independent variable X and student achievement would be the dependent variable Y. Student's attitudes and behavior are hypothesized to be the two mediators, the M variables.

Statistical approaches to the analysis of mediation have been discussed for many years in psychological literature. As early as 1948, Kenneth MacCorquodale and Paul E. Meehl talked about the value and logical status of so-called "intervening variables". Eagly and Chaiken (1993) showed that the effects of cognitive priming on attitude change were mediated by the accessibility of certain beliefs. Mediator model is also applied in organizational research (e.g., Makrs, Xaccaro & Mathieu, 2000; Mathieu, Heffner, Goodwin, Salas & CannonBowers, 2000), clinical research (e.g., Nolen Hoeksema & Jackson, 2001) and prevention research (MacKinnon & Dwyer, 1993). Baron & Kenny (1986), Sobel (1982), Alwin and Hauser and other researchers presented several different methods that can be used to assess a mediator model. MacKinnon and Hoffman also compared these different procedures for testing mediation to provide a general overview on for selecting the best test of mediation.

In this paper, firstly, I will attempt to use a generated dataset to build a multiple mediator model. Secondly, I will assess the model using the basic casual steps procedure, popularized by Baron and Kenny (1986). Next, I will test each mediated effect by calculating the associated confidence intervals. Then, in order to assess the difference among multiple mediators, I will apply three different widely used procedures (i.e., model with zero covariance between mediators; model with none-zero covariance between mediators; bootstrapping) to. Finally, I compare the results obtained from the different procedures and discuss the results.

METHODOLOGY

SINGLE MEDIATOR MODEL

Basic concepts

Mediation processes involving only one mediating variable is called simple mediation, also called single mediation. Figure 2 shows a model relating an independent variable X to a dependent variable Y, which has no mediator involved. Figure 3 describes such a process, which shows how the independent variable X's causal effect can be decomposed to its indirect effect ab on response variable Y through mediator M, as well as its direct effect c'.

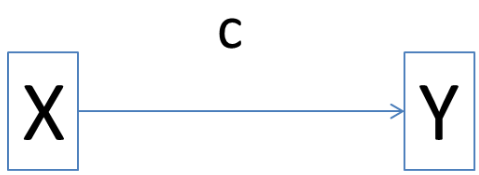

Figure 2: *Illustration of direct effect from X to Y.*

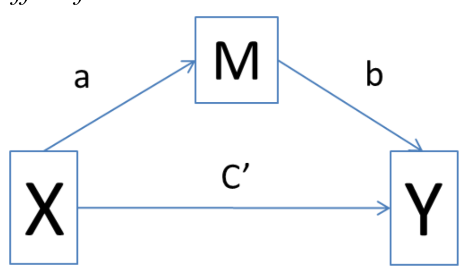

Figure 3: *Illustration of mediated effect from X to Y through M.*

The paths above can be tested by estimating each of the three equations below:

$$Y = i_1 + cX + \varepsilon_1 \tag{1}$$

$$Y = i_2 + c' + bM + \varepsilon_2 \tag{2}$$

$$M = i_3 + aX + \varepsilon_3 \tag{3}$$

where X is the independent variable, Y is the dependent variable and M is the mediating variable. c represents the total effect of X on Y, while c' represents the adjusted relation between X and Y after taking M into consideration (known as the direct effect). b is the parameter relating M to Y adjusted for the effects of X. a is the parameter relating X to M. ε_1, ε_2 and ε_3 are unexplained error terms. i_1, i_2 and i_3 are the relevant Equation's intercepts.

There are two approaches to measuring the mediation effect using the regression models in Equations (1) through (3). The most obvious approach to finding a mediation effect is by calculating the difference between c and c', i.e., $c - c'$. Since c' is the remaining direct effect after controlling for M, the difference between c and c' should be the effect caused by mediator. The other approach is to use the product of a and b as the mediated effect, because if substituting M in Equation (2) with Equation (3), the coefficient associated with M would be ab. Thus, $c - c' = ab$, which, on the other hand, mathematically proves that the total effect c can be decomposed into a direct effect c' and indirect effect (also called mediated effect) ab. ab can be interpreted as how much one unit change in X affects Y indirectly through M.

Estimation and Test Procedure

The parameters from Equation (1), (2), (3) can be estimated using Ordinary Least Squares Regression or Maximum Likelihood regression. The estimated values $\hat{a}$, $\hat{b}$, $\hat{c}$ and $\widehat{c'}$ will always fit the equation $\hat{c} - \hat{c}' = \hat{a}\hat{b}$, unless the sample size is different in different regression models. Once support is found for mediation, researchers are most interested in testing the statistical significance of the mediated effect. There are several methods for testing the mediated hypothesis. The most commonly used tests in educational and social sciences include Barron and Kenny's Causal Steps, and Product of Coefficients Test and Bootstrapping.

Causal Steps Strategy

The Causal Steps Strategy was popularized by Baron and Kenny (1986) in which they use three criteria. First X must be significantly related to Y. Next, X must be significantly related to M. Last but not the least, M must account for a significant amount of variability in Y after controlling for X. Note, however, that some researchers like Kenny, Kashy, and bolger (1998) argue that the first step (i.e., that there must be a significant total effect of X on Y) is not necessary for supporting mediation.

Product of Coefficients Test

Another general method to test mediation is to use the product of a and b and the product's standard error to perform a test of the significance of the mediation effect. The most commonly used standard error was derived by Sobel in 1982 using the delta method based on a first order Taylor series approximation. Thus, this test is also called the Sobel test.

The steps behind use of the Delta Method include: Suppose X is a random variable with EX = u. If we want to estimate a function g(x), a first order approximation based on a Taylor series expansion of g(x) around u would give us g(X)=g(u)+g'(u)(X-u). We can say that approximately,

$$\mathrm{Var}(\mathrm{g}(\mathrm{X})) = \left({\partial \mathrm{g}(\mathrm{X})}/{\partial \mathrm{X}}\right)^2 (\mathrm{Var}(\mathrm{X})) \tag{4}$$

For the variance of the product of $\hat{a}$ and $\hat{b}$, $\hat{a}\hat{b}$, the partial derivative with respect to $\hat{b}$ is $\hat{a}$ and the partial derivative with respect to $\hat{a}$ is $\hat{b}$. So the vector of partial derivatives D=[$\hat{b}, \hat{a}$] and the covariance matrix among $\hat{a}$ and $\hat{b}$ is

$V = \begin{bmatrix} s_{\hat{a}}^2 & Cov(\hat{a},\hat{b}) \\ Cov(\hat{a},\hat{b}) & s_{\hat{b}}^2 \end{bmatrix}$. Applying the formula gives the multivariate delta solution in Equation as derived by Sobel:

$$s_{\hat{a}\hat{b}}^2 = \hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2 + 2\hat{a}\hat{b}Cov(\hat{a},\hat{b}) \qquad (5)$$

It is commonly assumed that there is a zero covariance between $\hat{b}$ and $\hat{a}$, then the formula in Equation (5) is simplified to be: $s_{\hat{a}\hat{b}}^2 = \hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2$.

Similarly, after using the multivariate delta method, the standard error of the mediated effect in terms of difference between c and c' can be written like

$$s_{\hat{c}-\widehat{c'}} = \sqrt{s_{\hat{c}}^2 + s_{\widehat{c'}}^2 - 2s_{\hat{c}\widehat{c'}}} \qquad (6)$$

Bootstrapping

The limitation of using the Delta Method to derive standard error is that its use requires an assumption of multivariate normality for the sampling distribution of $\hat{a}\hat{b}$. This assumption is reasonable for a dataset with a sufficiently large sample size. However, in a considerable number of studies, the sample size is likely not sufficiently large. In 2002, Shrout and Bolger suggested using bootstrapping to solve this problem.

To bootstrap, repeatedly take a sample of size n, k times with replacement from the original sample. Estimate $\hat{a}\hat{b}$ for each bootstrapped sample. If k is large enough, say 5,000, the resulting distribution of $\hat{a}\hat{b}$ s can serve as empirical, nonparametric approximations of the sampling distribution of $\hat{a}\hat{b}$. In such a way can we see whether the mediated effect is statistically significant or not, or what might be a reasonable range for the effect rather than a single value. This is accomplished by sorting these k values from smallest to largest. In the ordered dataset, the 95% confidence interval would be the interval between the 2.5th percentile and 97.5th percentile values. If zero is not in this interval, we can claim that the indirect effect $\hat{a}\hat{b}$ is significantly different from zero. Otherwise, the indirect effect is not significant.

MULTIPLE MEDIATOR MODEL

Basic Concept

Often researchers have more than one mediator that might account for an X-Y relationship, as in the example mentioned at the beginning. Thus, we need a multiple mediator model. For the same process in Figure 2, Figure 4 depicts both the direct effect of X on Y and the indirect effect from multiple mediators.

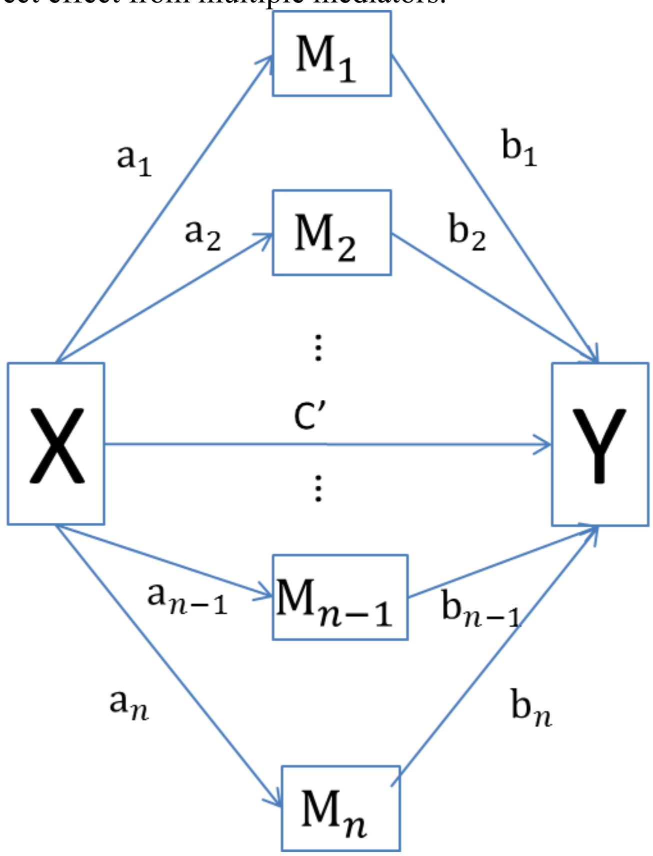

Figure 4: *Illustration of a multiple design with n mediators.*

Since this paper focuses on models with three mediators, all the illustrations and calculations will be based on a model with three mediators. Similar to the single mediator model, there are five equations describing the paths above with five mediators.

$$Y = i_1 + cX + \varepsilon_1 \tag{7}$$

$$Y = i_2 + c'X + b_1M_1 + b_2M_2 + b_3M_3 + \varepsilon_2 \tag{8}$$

$$M_1 = i_3 + a_1X + \varepsilon_3 \tag{9}$$

$$M_2 = i_4 + a_2X + \varepsilon_4 \tag{10}$$

$$M_3 = i_5 + a_3X + \varepsilon_5 \tag{11}$$

Where X is the independent variable, Y is the dependent variable, M_1, M_2 and M_3 are first, second and third mediators, respectively. c is called the total effect and c' is called the direct effect. The specific indirect effect of X on Y through a mediator M_i (i=1,2,3) is defined as the product of a_i and b_i. The total indirect effect is the summation of each single indirect effect, which is $\sum_{i=1}^{3} a_i b_i$. The total effect then can be decomposed into direct effect and indirect effects, i.e. $c = c' + \sum_{i=1}^{3} a_i b_i$. The total indirect effect can also be written as c-c'.

Estimation and Test Procedure

The estimation procedure is basically the same as for the single mediator model, using Ordinary Least Squares Regression or Maximum Likelihood regression.

Causal Steps Strategy

Causal step method mentioned above is still applied here with a few limitations. The basic steps are

(1) X must affect Y. c must be significant in Equation (7)

(2) X must affect each of the mediators M_1, M_2 and M_3. a_1, a_2, a_3 must then be statistically significant in Equations (9)-(11)

(3) Each mediator must affect Y after controlling for X. b_1, b_2, b_3 must each be statistically significant in Equation (8)

(4) For evidence supporting full mediation, the direct effect from X must be nonsignificant [i.e., c' would be not statistically significant in Equation (8)].

Difference in Coefficient Test

While testing multiple mediation hypotheses is different and is easier when tested in terms of c-c’, because using $\sum_{i=1}^{3} a_i\, b_i$ involve more calculation. The standard error of total mediated effect is shown in Equation (12).

$$s_{\hat{c}-\widehat{c\prime}} = \sqrt{s_{\hat{c}}^2 + s_{\widehat{c\prime}}^2 - 2rs_{\hat{c}}s_{\widehat{c\prime}}} \tag{12}$$

where $rs_{\hat{c}}s_{\widehat{c\prime}}$ is the covariance between $\hat{c}$ and $\widehat{c'}$. However use of the total mediated effect does not permit assessment of the mediated effect for *each* mediator.

Contrast

Because we are now focusing on a model with multiple mediators, we also need to consider contrast hypotheses intended to assess potential differences between pairs of mediated effects. This would be more complex to achieve, because assessment may involve simultaneous testing of multiple mediation effects and correlation between mediators.

Consider the situation where we want to contrast the indirect effect via M_1 versus that through M_2. The estimated value of the contrast is calculated by Equation (13).

$$\Delta_{M_1-M_2} = a_1b_1 - a_2b_2 \tag{13}$$

According to Sobel, use of the delta method yields the following formula for the standard error of the difference statistic (in Equation 13):

$$S_{\Delta_{M_1-M_2}} = \sqrt{b_1^2s_{a_1}^2 - 2b_1b_2s_{a_1,a_2} + b_2^2s_{a_2}^2 + a_1^2s_{b_1}^2 - 2a_1a_2s_{b_1,b_2} + a_2^2s_{b_2}^2} \tag{14}$$

The result is the standard error of the contrast and can be used to test this comparison hypothesis or to construct a confidence interval around the estimate of the difference in the two mediators’ indirect effects.

In the multiple mediation model, for example, if we have three mediators and thus three pairwise contrasts and we want to test three hypotheses or construct three confidence interval simultaneously each at a certain significance level, say α, the probability of having one or more type I errors across the set of three tests will be more than $1-(1-\alpha)^3$. If α is 0.05, this number will be 0.14, which is does not provide a reasonable experiment-wise Type I error rate. In order to address this inflated experiment-wise Type I error problem, we could use Bonnferroni's adjustment to modify the significance level used when constructing the relevant confidence intervals and thus the resulting critical value. To obtain the experiment-wise type I error rate of α for a set of *g* contrasts, we would use α/g as the per-comparison significance level.

Bootstrapping

As mentioned in the presentation of the single mediator model, bootstrapping can also be used here to assess the statistical significance of the indirect effects or of their contrasts. Bootstrapping is a very general approach and it can be used to make inferences about any parameters we are interested in.

METHODS

Suppose we are interested in the effects of teaching methods on student achievement. We are curious whether such an effect can be mediated through students' confidence level, motivation level and cooperation level. So we attempt to use some assumption to generate a dataset and fit a multiple mediator model to solve this question.

The dataset contains five variables: X, M1, M2, M3 and Y. 1,000 values for each variable were generated from a multivariate normal distribution, with means of 50 for each variable and respective variance of 3,12,14,13 and 20. The correlation matrix used for generating the dataset is shown below, and correlation matrix was calculated by path tracing.

correlation matrix					
	x	M1	M2	M3	y
X	1				
M1	0.56	1			
M2	0.49	0.2744	1		
M3	0.6	0.336	0.294	1	
Y	0.7054	0.69704	0.573616	0.75604	1

Table 1: *Correlation matrix among independent variable, dependent variable and mediators.*

Assume that X represents teacher experience and Y represents student achievement. The three mediators are students' confidence level (M1), motivation level (M2) and cooperation level (M3). The path model is illustrated below.

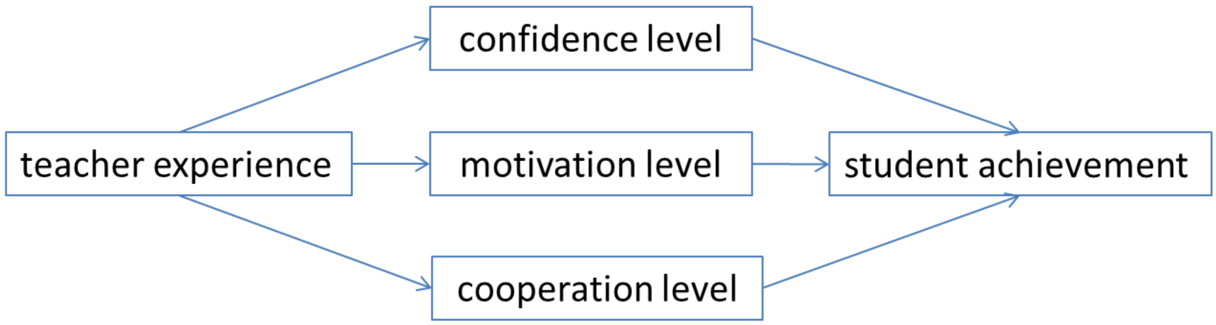

Figure 5: *Example of case used in this paper.*

Statistical software RStudio 0.97.312 was used in this paper. Appendix contains codes used in the experiment in R.

EXPERIMENTAL RESULTS

From *lm* regression output in R, we have the estimated multiple mediator models presented below.

When estimating the total effect using Equation (7), the following parameter estimates resulted: $\widehat{\mathrm{Y}} = 20.0321 + 4.7984\mathrm{X}$. The standard error of c was 0.1472. Next, the full model depicted in Figure 5 was estimated with following results: $\widehat{\mathrm{Y}} = 4.4984 - 0.0341\mathrm{X} + 0.7074\mathrm{M}_1 + 0.4144\mathrm{M}_2 + 0.8019\mathrm{M}_3$. The standard errors of c', b_1, b_2 and b_3 were 0.1149, 0.0221, 0.0175 and 0.0207 respectively. To obtain each mediator's value for a, the formulas from Equation (9) to Equation (11) were estimated for each mediator with the following results, for M1, $\widehat{\mathrm{M}}_1 = 0.2982 + 2.4697\mathrm{X}$, and the standard error of a_1 was 0.1056; for M2, $\widehat{\mathrm{M}}_2 = 7.7878 + 2.464\mathrm{X}$ and the standard error of a_2 was 0.1329; for M3, $\widehat{\mathrm{M}}_3 = 15.0875 + 2.5743\mathrm{X}$ and the standard error of a_3 was 0.1126.

Teacher experience is significantly related to student achievement, providing strong evidence that there is a statistical significant relation between the independent variable and dependent variable. And this total effect can be explained by teacher experience influences student's confidence level (M1), motivation level (M2) and cooperation level (M3), because

- There was a significant effect of teacher's experience on student's confidence level (M1) ($\hat{a}_1$=2.4697, $S_{\hat{a}_1}$=0.1056, $p<0.005$), as well as on student's motivation level (M2) ($\hat{a}_2$=2.464, $S_{\hat{a}_2}$=0.1329, $p<0.005$) and student's cooperation level (M3) ($\hat{a}_3$=2.5743, $S_{\hat{a}_3}$=0.1126, $p<0.005$)
- these three mediators significantly accounts for variability in Y when controlling for X ($\hat{b}_1$=0.7074, $S_{\hat{b}_1}$=0.0221, $p<0.005$; $\hat{b}_2$=0.4144, $S_{\hat{b}_2}$=0.0175, $p<0.005$; $\hat{b}_3$=0.8019, $S_{\hat{b}_3}$=0.0207, $p<0.005$)

- the effect of X on Y decreases substantially when mediators entered simultaneously when X as a predictor as Y. The drop is 4.8325 in the value of $\hat{c}'$ (-0.0341) compared with $\hat{c}$ (4.7984).

SPECIFIC MEDIATED EFFECT

Product of Coefficients

The three mediated effect are estimated as below:

Mediation through confidence level: $\hat{a}_1\hat{b}_1$=(2.4697)(0.7074)=1.7471

Mediation through motivation level: $\hat{a}_2\hat{b}_2$=(2.464)(0.4144)=1.0211

Mediation through cooperation level: $\hat{a}_3\hat{b}_3$=(2.5743)(0.8019)=2.0643

The total mediation effect is $\hat{a}_1\hat{b}_1 + \hat{a}_2\hat{b}_2 + \hat{a}_3\hat{b}_3$ = 1.7471+1.0211+2.0643=4.8325, which is the same as $\hat{c} - \hat{c}'$=4.7984-(-0.0341).

Using Equation (5), the standard error of a specific mediated effect $\hat{a}\hat{b}$ is

$$S_{\hat{a}_1\hat{b}_1} = \sqrt{(2.4697)^2(0.0221)^2 + (0.7074)^2(0.1056)^2} = 0.0925$$

$$S_{\hat{a}_2\hat{b}_2} = \sqrt{(2.464)^2(0.0175)^2 + (0.4144)^2(0.1329)^2} = 0.0699$$

$$S_{\hat{a}_3\hat{b}_3} = \sqrt{(2.5743)^2(0.0207)^2 + (0.8019)^2(0.1126)^2} = 0.1048$$

The estimated mediation effects and their corresponding standard errors yield the z statistics, which are $z_{\hat{a}_1\hat{b}_1} = \frac{\hat{a}_1\hat{b}_1}{S_{\hat{a}_1\hat{b}_1}} = \frac{1.7471}{0.0925} = 18.8876$, $z_{\hat{a}_2\hat{b}_2} = \frac{\hat{a}_2\hat{b}_2}{S_{\hat{a}_2\hat{b}_2}} = \frac{1.0211}{0.0699} = 14.608$ and $z_{\hat{a}_3\hat{b}_3} = \frac{\hat{a}_3\hat{b}_3}{S_{\hat{a}_3\hat{b}_3}} = \frac{2.0643}{0.1048} = 19.6975$. If we want to test whether they are all significant at 95% confidence level, we have to use Bonnferroni adjustment. Since $z_{0.05*/2*3}$ =2.39 and all the sample test statistics are much greater than the critical *z*-score, we can say that the three mediation effects from student's confidence level, motivation level and cooperation level are each statistically significant ($p < .05$).

Also, the 95% confidence interval of mediated effect through students' confidence level, motivation level and cooperation level are (1.526, 1.9682), (0.854, 1.1882) and (1.8138, 2.3148), respectively.

Bootstrapping

When we have a dataset with large number of records, say n, there is a reason to choose $n/\sqrt{2}$ as sample size and use this sample to do analysis. Thus, in our case, the sample size should be greater than 707. In order to obtain more accuracy, in bootstrapping, I chose to take 5,000 samples of 900 cases with replacement from the original sample and calculated each mediated effect. Appendix includes the R code with *sample* command of Bootstrapping.

The sampling distribution of mediated effect from confidence level is showed in Figure 6.

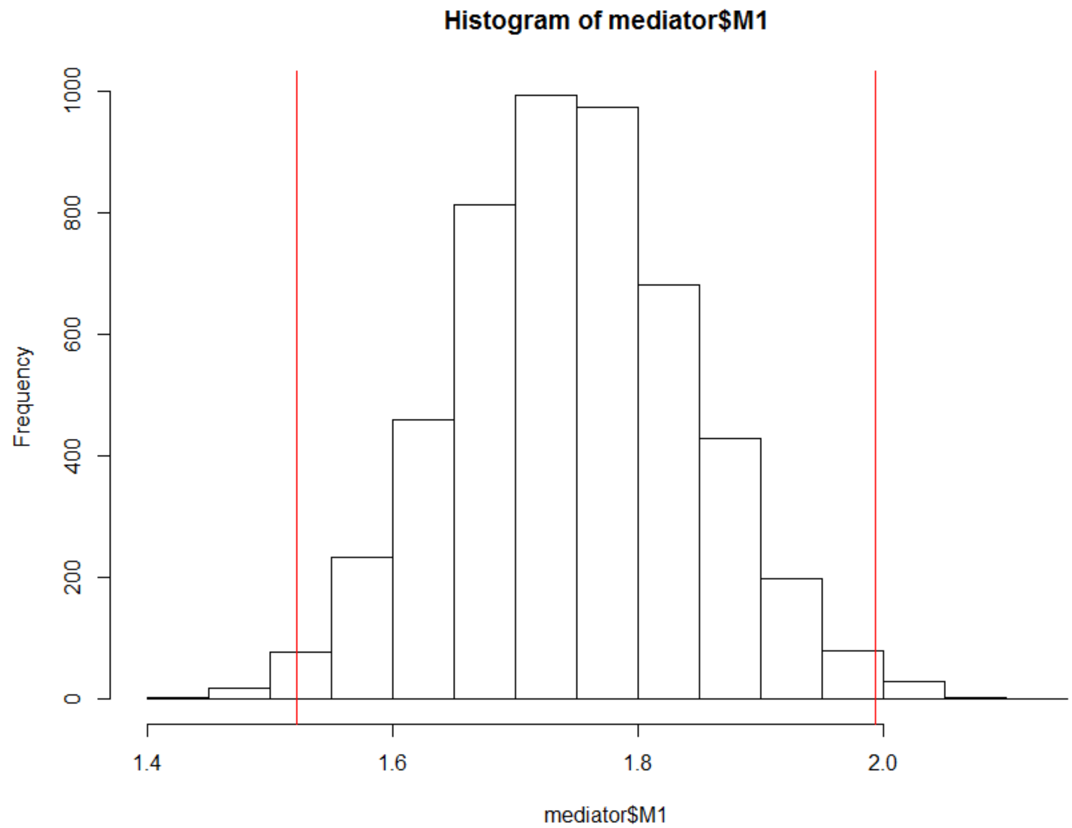

Figure 6: *Histogram of mediated effect through confidence level. The red lines represent lower bound and upper bound of 95% confidence interval.*

To maintain the Type I error, still, we applied the Bonnferroni procedure, so $\alpha = 0.05/3 = 0.0167$ and the confidence interval limits comprised the 0.83th percentile and 99.17th percentile values.

Similar procedure for mediated effect from motivation level and cooperation level, the sampling distributions are shown below.

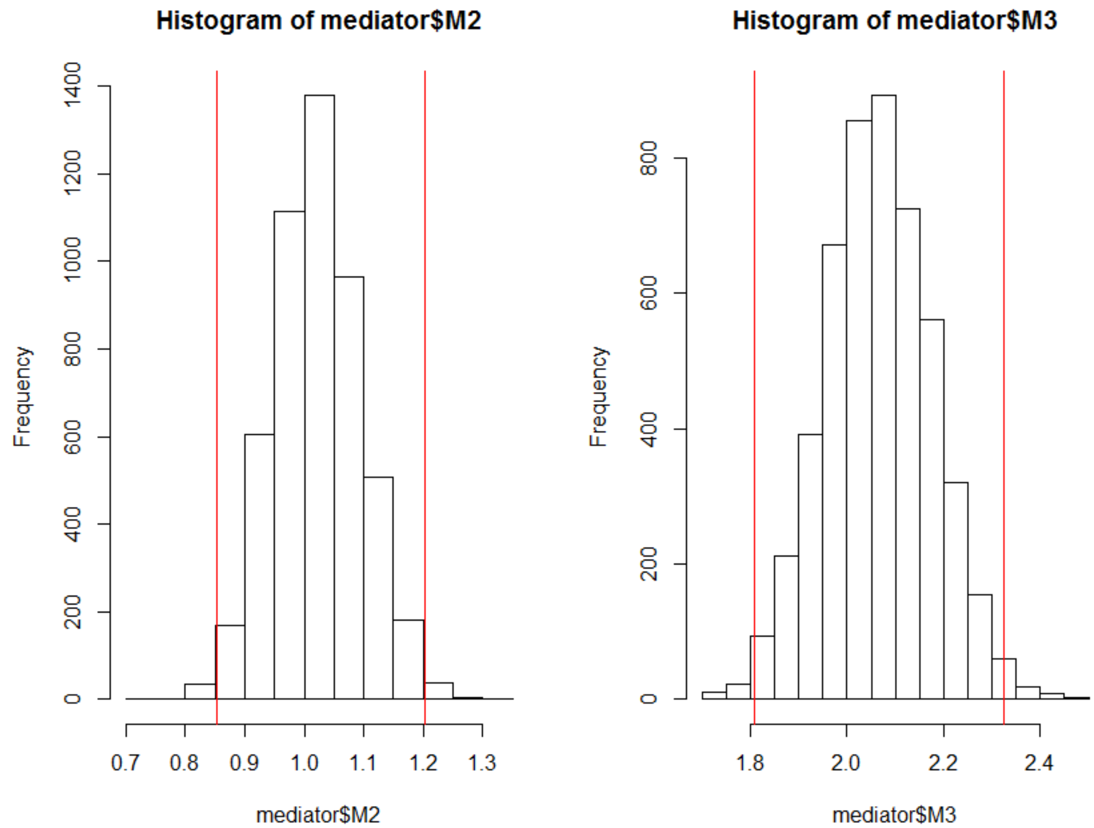

Figure 7: *Histogram of mediated effect through motivation level and cooperation level. The red lines represent lower and upper bounds of each 95% of confidence interval.*

So using Bonnferroni's correction, the 95% confidence intervals for the mediated effects through the confidence level, motivation level and cooperation level variables were (1.5219, 1.9935), (0.8518, 1.203) and (1.8083, 2.3239), respectively. None of these confidence intervals contained zero and thus, we can say that each of these three mediated effects was statistically significant.

CONTRAST

Zero Covariance

Using Equation (13) and (14), the difference between students' confidence level and motivation level is

$$\Delta_{M_1-M_2} = \hat{a}_1\hat{b}_1 - \hat{a}_2\hat{b}_2 = 1.7471 - 1.0211 = 0.726$$

$$S_{\Delta_{M_1-M_2}} = \sqrt{b_1^2 s_{a_1}^2 + b_2^2 s_{a_2}^2 + a_1^2 s_{b_1}^2 - 2a_1a_2 s_{b_1,b_2} + a_2^2 s_{b_2}^2}$$

$$= \sqrt{(0.0925)^2 + (0.0699)^2 - 2(2.4697)(2.464)(-0.000015)} = 0.1167$$

$$z_{M_1-M_2} = \frac{\Delta_{M_1-M_2}}{S_{\Delta_{M_1-M_2}}} = \frac{0.726}{0.1167} = 6.221$$

Similar, the difference between the mediated effects through confidence level and cooperation level is

$$\Delta_{M_1-M_3} = \hat{a}_1\hat{b}_1 - \hat{a}_3\hat{b}_3 = 1.7471 - 2.0643 = -0.3172$$

$$S_{\Delta_{M_1-M_3}} = \sqrt{b_1^2 s_{a_1}^2 + b_3^2 s_{a_3}^2 + a_1^2 s_{b_1}^2 - 2a_1a_3 s_{b_1,b_3} + a_3^2 s_{b_3}^2}$$

$$= \sqrt{(0.0925)^2 + (0.1048)^2 - 2(2.4697)(2.5743)(0.000025)} = 0.3362$$

$$z_{M_1-M_3} = \frac{\Delta_{M_1-M_3}}{S_{\Delta_{M_1-M_3}}} = \frac{|-0.3172|}{0.3362} = 0.9435$$

The difference between the mediated effects through motivation level and cooperation level is

$$\Delta_{M_2-M_3} = \hat{a}_2\hat{b}_2 - \hat{a}_3\hat{b}_3 = 1.0211 - 2.0643 = -1.0432$$

$$S_{\Delta_{M_2-M_3}} = \sqrt{b_2^2 s_{a_2}^2 + b_3^2 s_{a_3}^2 + a_2^2 s_{b_2}^2 - 2a_2a_3 s_{b_2,b_3} + a_3^2 s_{b_3}^2}$$

$$= \sqrt{(0.0699)^2 + (0.1048)^2 - 2(2.5743)(2.464)(0.00000127)} = 0.0159$$

$$z_{M_2-M_3} = \frac{\Delta_{M_1-M_2}}{S_{\Delta_{M_1-M_2}}} = \frac{|-1.0432|}{0.0159} = 65.61$$

After applying Bonnferroni adjustment, $z_{\alpha/2} = z_{0.05*/2*3} = 2.39$ while only $z_{M_1-M_3}$ is smaller than 1.44, we can say that the at 95% confidence level, only the difference between the mediated effects through confidence level versus cooperation level were not statistically significant while the difference between the effects through confidence level versus motivation level and the difference between the effects through motivation level and cooperation level were statistically significant.

Non-zero Covariance

In order to get the covariance between parameters from different regression equations, indicators were involved to address this problem. Let sy1, sm1, sm2, sm3 be the indicators for Equation 8, 9, 10, 11 respectively. The R output provides the covariance matrix among a_1, a_2 and a_3 as Table 2.

	a_1	a_2	a_3
a_1	1.172645e-02	-8.688699e-34	5.809257e-34
a_2	-8.688699e-34	1.172645e-02	3.442831e-33
a_3	5.809257e-34	3.442831e-33	1.172645e-02

Table 2: *Covariance matrix among coefficient associated with each mediator.*

We can see that even the correlation between different mediators are not weak (as shown in Table 1), the covariance between their coefficients with independent variable is almost zero. So even if we consider the possibility of a non-zero covariance in calculation of the test statistic's standard error, the results will not change that much. Thus, in this case, it's reasonable to assume zero-covariance between those coefficients.

Bootstrapping

Although we get almost same results from with and without assumption of zero-covariance, there is still a big limitation that the use of Delta Method requires an assumption of multivariate normality for the sampling distribution of estimates of the ab parameter.

The sampling distributions from Boostrapping are shown below.

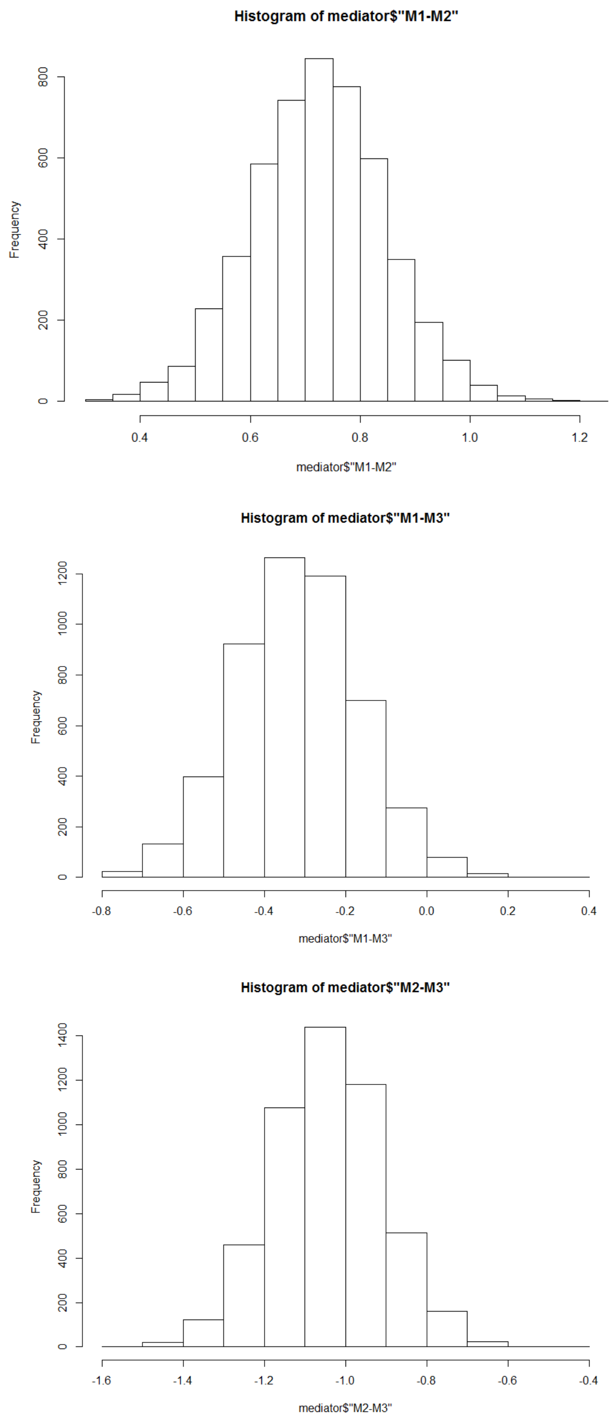

Figure 8: *Histograms of sampling distribution of difference between mediated effect from confidence level & motivation level, confidence level & cooperation level and motivation level & cooperation respectively.*

So at 95% confidence level, the difference between mediated effects from confidence level and motivation level, the difference between confidence level and cooperation level and the difference between motivation level and cooperation level will fall in (0.4299, 1.0262), (-0.6769, -0.0453), and (-1.3694, -0.7155) respectively. As we can see from these three confidence intervals, none of them contains zero. Based on the relationship between hypothesis test and confidence interval, we can say that *all* of the three pairwise differences in the mediated effects were significant at 95% confidence level, which is a big difference from the results we got before.

CONCLUSION

Mediator Model has been devoted to seeking the mechanism behind causal effect of two variables in recent years and it's widely used in psychology, education and may other fields. As more and more researchers become interested in testing mediator hypotheses in their research, there are more and more methods that can be used to assess mediator models. Each method has its own assumptions, strengths and weaknesses. In the experiment presented in this paper, there are two major findings. One is that unlike before, when most researchers assumed the covariance between coefficients from different equations to zero, I used indicator model to get the covariance and include it in the calculation. Although there was not a big difference between two procedures, it proved, on the other hand, that it might be reasonable to assume that this covariance is zero in future research. In addition, I found that bootstrapping might be more appropriate than the causal steps approach, as well as the Sobel test, because bootstrapping does not require distributional assumptions. And it gave different results.

There is still much to do on the comparison between different methods. For example, the confidence interval from the general procedure can be narrower than that from Bootstrapping, which should be better because it means more precision. But it has also been proved that the sampling distribution of estimates of ab tends to be asymmetric with nonzero skewness and kertosis (Bollen & Stine, 1990)[5]. In addition, the correlation between mediators is not weak, but it's also not very strong. And this may be the cause of small covariance. Thus, future work may be needed to address these problems.

Appendix

R code

```
library(MASS)
###generate dataset###
Sigma <-
matrix(c(9,20.16,20.58,23.4,42.324,20.16,144,46.0992,52.416,167.2896,20.58,46.0992,1
96,53.508,160.6125,23.4,52.416,53.508,169,196.5704,42.324,167.2896,160.6125,196.57
04,400),5,5)
Sigma
data<-mvrnorm(n=1000, c(10,25,32,41,68), Sigma)
data<-as.data.frame(data)
colnames(data) <- c("X", "M1","M2","M3","Y")
###fit multiple mediation model###
summary(lm(Y~X+M1+M2+M3, data=data))
summary(lm(Y~X, data=data))
summary(lm(M1~X, data=data))
summary(lm(M2~X, data=data))
summary(lm(M3~X, data=data))
vcov(lm(Y~X+M1+M2+M3, data=data))
###use indicator model to find covariance###
library(reshape2)
data$fid <- 1:nrow(data)
stacked <- melt(data, id.vars = c("X","fid","M1","M2","M3","Y"), measure.vars =
c("Y","M1","M2","M3"), value.name = "Z")
stacked <- within(stacked, {
sy1 <- as.integer(variable == "Y")
sy2 <- as.integer(variable == "Y")
sm1 <- as.integer(variable == "M1")
sm2 <- as.integer(variable == "M2")
sm3 <- as.integer(variable == "M3")
})
library(nlme)
mixed<-lm(Z~-
1+sm1+sm1:X+sm2+sm2:X+sm3+sm3:X+sy1+sy1:X+sy1:M1+sy1:M2+sy1:M3,
data=stacked)
summary(mixed)
vcov(mixed)
###bootstrap
rep<-5000
size<-900
mediator<-matrix(0,ncol=6,nrow=rep)
sample<-matrix(0,ncol=6,nrow= size)
```

```
mediator <- data.frame(mediator)
sample <- data.frame(sample)
for (i in 1:rep){
id<-sample(c(1:1000), size, replace = TRUE)
  for (j in 1: size){
    n<-match(id[j],data$fid)
    sample[j,]<-data[n,]
    colnames(sample) <- c("X", "M1","M2","M3","Y","id")
  }
reg1<-lm(Y~X+M1+M2+M3, data=sample)
reg2<-lm(Y~X, data=sample)
reg3<-lm(M1~X, data=sample)
reg4<-lm(M2~X, data=sample)
reg5<-lm(M3~X, data=data)
a1<-coefficients(reg3)[2]
a2<-coefficients(reg4)[2]
a3<-coefficients(reg5)[2]
b1<-coefficients(reg1)[3]
b2<-coefficients(reg1)[4]
b3<-coefficients(reg1)[5]
m1<-a1*b1
m2<-a2*b2
m3<-a3*b3
mediator[i,]<-c(m1,m2,m3,m1-m2,m1-m3,m2-m3)
}
colnames(mediator) <- c("M1","M2","M3","M1-M2","M1-M3","M2-M3")
par(mfrow=c(1,1))
qts <- quantile(mediator[,1],probs=c c(.0083,.9917))
hist(mediator$M1)
abline(v=qts[1],col="red")
abline(v=qts[2],col="red")
par(mfrow=c(1,2))
qts <- quantile(mediator[,2],probs=c(.0083,.9917))
hist(mediator$M2)
abline(v=qts[1],col="red")
abline(v=qts[2],col="red")
qts <- quantile(mediator[,3],probs=c(.0083,.9917))
hist(mediator$M3)
abline(v=qts[1],col="red")
abline(v=qts[2],col="red")
par(mfrow=c(3,1))
hist(mediator$'M1-M2')
hist(mediator$'M1-M3')
```

```
hist(mediator$'M2-M3')
quantile(mediator[,4],probs= c(.0083,.9917))
quantile(mediator[,5],probs= c(.0083,.9917))
quantile(mediator[,6],probs= c(.0083,.9917))
```

Reference

1. Kristopher J. Preacher, Andrew F. Hays. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods 2008; **40**(3): 879-891.

2. Kristopher J. Preacher, Andrew F. Hays. SPSS and SAS procedure for estimating indirect effects in simple mediation models. Behavior Research Methods, Instruments & Computers 2004; **36**(4): 717-731.

3. David P. MacKinnon, Chondra M. Lockwood, Jeanne M. Hoffman, Stephen G. West, Virgil Sheets. A Comparison of Methods to Test Mediation and Other Intervening Variable Effects 2002; **7**(1): 84-104.

4. Patrick E. Shrout, Niall Bolger. Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations. Psychological Methods 2002; **7**(4): 422-445.

5. Andrew F. Hayes. Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. Communication Monographs 2009; **76**(4): 408-420.

6. David P. Mackinnon. Introduction to Statistical Mediation Analysis. Lawrence Erlbaum Associates 2011.